ARI Research Note 92-79

Evaluation of Relative Importance Judgment Methods in the Context of Causal Prediction

Robert M. Hamm
U.S. Army Research Institute

Field Unit at Fort Leavenworth, Kansas
Stanley M. Halpin, Chief

Manpower and Personnel Research Division
Zita M. Simutis, Director

August 1992

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

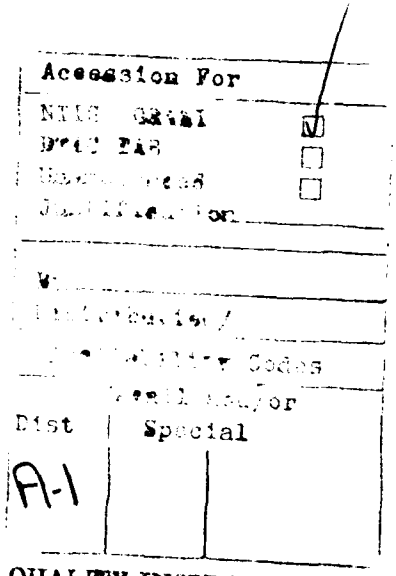
A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

MICHAEL D. SHALER
COL, AR
Commanding

Technical review by

James W. Lussier
Douglas K. Spiegel



REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1992, August	Interim Oct 88 - Jan 91

4. TITLE AND SUBTITLE

Evaluation of Relative Importance Judgment Methods in the Context of Causal Prediction

5. FUNDING NUMBERS

61102B
74F
4901
C01

6. AUTHOR(S)

Hamm, Robert M.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-RH
P.O. Box 3407
Fort Leavenworth, KS 66027-0347

8. PERFORMING ORGANIZATION REPORT NUMBER

ARI Research Note 92-79

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

--

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

--

11. SUPPLEMENTARY NOTES

Author was a National Research Council Senior Associate working in the ARI Fort Leavenworth Field Unit, 1988-1991.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

--

13. ABSTRACT *(Maximum 200 words)*

This study investigated five methods for measuring the relative importance of situational factors in predicting mission success in a planning context. The methods were ranking, 2-category rating (yes or no), ratio scale rating (magnitude estimation), probability change, and necessity or sufficiency judgments. Interpersonal agreement was highest using the probability change method. Comparison of the importance profiles produced by the various methods showed that ranking, rating, and yes or no were most similar to each other. Respondents indicated their preference among the methods and evaluated the methods' usefulness for a number of purposes. The ranking and probability change methods were most preferred and rated most useful.

14. SUBJECT TERMS

Subjective probability
Relative importance
Measurement
Judgment
Prediction
Situation assessment

15. NUMBER OF PAGES

31

16. PRICE CODE

--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

This research note represents the application of judgment and decision-making techniques to military officers' understanding of factors in the battlefield situation pertinent to the success of planned courses of action. The need for methods to facilitate judgment of relative importance and communication of these judgments is widespread in operational and research contexts. This work was carried out in the Fort Leavenworth Field Unit of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) MANPRINT Division, whose mission is improvement of the efficiency, accuracy, and timeliness of command and control. The work was performed under the auspices of the National Research Council Associateship program.

ACKNOWLEDGMENT

The author's tenure at the U.S. Army Research Institute for the Behavioral and Social Sciences Fort Leavenworth Field Unit was supported by a Senior Research Associate award from the National Research Council. Thanks are due to the armchair tacticians who helped create the two problems, especially Lloyd Crumley and Rex Michel. Ernie Lowden, Lieutenant Colonel William Brethorst, and Captain Mike Spry helped arrange subjects' participation. Reid Hastie made suggestions about the organization of the manuscript, and four reviewers proposed useful changes.

Portions of this work were presented at the Seventh Annual Workshop on Command and Control Decision Aiding, Dayton, Ohio, April 1990.

EVALUATION OF RELATIVE IMPORTANCE JUDGMENT METHODS
IN THE CONTEXT OF CAUSAL PREDICTION

EXECUTIVE SUMMARY

Requirement:

In several contexts in the command and control of military forces, there is a need for subjective judgment concerning the relative importance of different factors in a situation. But little research has been done to investigate the meaningfulness, accuracy, or reliability of such judgments. To address this lack of information, five alternative methods were compared.

Procedure:

The subjects of this research were 153 students of the Command and General Staff College. Each read a description of a problem situation in which a commander had to produce a mission plan. The commander's plan was also described. Two problems were used--one about a response to a crashed helicopter and the other about an attack across a river. For each problem, the subject estimated the mission's probability of success and rated the relative importance of eight situational factors.

Each subject used two of the following five methods to evaluate the importance of the factors: categorical yes or no judgments, numerical ranking, numerical rating, categorical necessity and sufficiency judgments, and numerical probability change judgments. Each method was rated with respect to seven issues: (1) ease of use, (2) confidence in accuracy of the method, (3) helpfulness for commander communicating to staff, (4) helpfulness for guiding the scheduling of planning tasks, (5) helpfulness for guiding the allocation of resources, (6) usefulness for explaining mission success, and (7) usefulness for explaining mission failure.

Findings:

Some subjects exhibited internal incoherence using the necessity or sufficiency and probability change methods. Subjects showed the greatest interpersonal agreement on the probability change method. Interpersonal agreement about the rating and ranking methods was not stable between problems. The judgments of which factors were important were similar using different methods. The range of importance scores between the most and least important factors was smallest with the yes or no method and largest with the rating method. The rating, ranking, and yes

or no methods produced importance scores that were fairly similar to each other; the probability change method and especially the necessity or sufficiency method produced scores that were less similar. The ranking and probability change methods were most preferred by the subjects and received the highest ratings on six of the seven issues. The yes or no method was considered easiest. The rating and necessity or sufficiency methods were least preferred.

Utilization of Findings:

The fact that different methods produce similar profiles of the factors' importance validates the general concept of relative importance. However, when methods have systematic disagreements (as do the probability change and necessity or sufficiency methods), the differences suggest that there is an important distinction between them; analysis is needed to determine which method to use in each application. Those methods that produce similar profiles can probably be substituted for each other according to user preference.

EVALUATION OF RELATIVE IMPORTANCE JUDGMENT METHODS
IN THE CONTEXT OF CAUSAL PREDICTION

CONTENTS

LIST OF TABLES

CONTENTS (Continued)

Page

LIST OF FIGURES

EVALUATION OF RELATIVE IMPORTANCE JUDGMENT METHODS IN THE CONTEXT OF CAUSAL PREDICTION

Introduction

In the command and control of military forces, there is need for subjective judgment concerning the relative importance of different factors in the situation. These judgments are sometimes explicit, as in the determination of "the most vital units for successful accomplishment of the mission" in order to "establish priorities for support" (U.S. Army, 1986, p. 70). More often, the judgments are implicit, as in war gaming, which is said to "highlight tasks that appear to be particularly important to the operation..." (CGSC, 1988, p. 4.1), but for which the doctrine does not specify what is to be done with those tasks identified as important.

Although expressions of relative importance occur in many situations, little is known about the ability of military officers to make such judgments or to communicate them to others. For operational purposes, it has been assumed that experienced officers can adequately assess the relative importance of information, goals, resources, and threats. Thus they are advised, "Since time is precious, establishing priorities helps subordinate leaders and commanders determine the order in which the tasks must be accomplished" (U.S. Army, 1987, p. 91), but they are not given specific guidance about what should be given priority. Further, they are judged on their ability to make relative importance judgments. Field Manual 101-5 states, "the judgment of a commander is indicated by his ability to select the important factors in any given situation and to accord them due weight" (U.S. Army, 1954, p 92; similar statements have occurred in subsequent revisions).

Research in command and control also depends on judgments of relative importance. It has been assumed that officers can respond appropriately when asked, for example, to

> "put... your opinion of how essential each [of the eight major elements of the order of battle] is according to the following scale of values, entering the applicable number in the column labeled 'Essentiality Rating.'
>
> 1 = Essential
> 2 = Important
> 3 = Useful
> 4 = Of some use
> 5 = Of little use"

(Coates and McCourt, 1976, p. 58). But little research has been done in a military context to establish the meaningfulness,

accuracy, and reliability of such judgments of relative importance. Judgment and decision making researchers have studied people's reports of the relative importance of attributes in their preference judgments (Goldstein and Beattie, in press; Reilly and Doherty, 1989; Shanteau, 1980), showing that in many situations they are inaccurate. Comparisons have been made between methods for expressing relative importance in the contexts of stating preferences (e.g., Cook and Stewart, 1975) and explaining one's own actions (Jaccard and Sheng, 1984). But in military command and control, relative importance concepts are required for a broader range of situations than preference and self explanation.

Measures of relative importance are also of interest to developers of computerized decision aids (e.g., Andriole, 1987). The potential exists for interaction between the formal analysis that a decision aid can provide and the staff officer's subjective judgments. Two specific modes of interaction are: (a) the officer produces judgments of relative importance and the decision aid translates these into a recommendation for action (e.g., a schedule); and (b) the decision aid takes available information and produces a measure of relative importance (e.g., an alarm).

The need for diverse methods for eliciting relative importance judgments

Relative importance judgments can be useful in a variety of contexts. An understanding of the function of such judgments in each context would make it possible to make maximum use of the judgments and to avoid confusion.

Analysis of the functions of relative importance judgments should focus on what difference they might make in the behavior of the commander or the staff. For example, it might be important to watch a vulnerable flank throughout an action so that one can respond promptly if it is threatened. It might be important to obtain information about enemy capabilities at the beginning of a planning session because everything else will depend on it. For the flank to be more important than the rear means "check the flank more frequently than the rear"; for the information about enemy capabilities to be more important than information about the terrain means "get enemy information before terrain information". Although the recipient of a message conveying relative importance will often know what meaning is intended from the context, there are times when the context is ambiguous and misinterpretation of the message is possible.

Five types of relative importance. There are at least five contexts which may require relative importance judgments. These are resource allocation, scheduling, prediction, information gathering, and preference. Each of these types of relative importance can be formally measured, but only when several attributes of the situation have been specified. These

attributes are: the person whose interests are judged, the goal or objective, the partition of the world that identifies relevant factors, and the behaviors that will be affected by the judgment of the relative importance of the factors (Hamm, 1990b). For example, the relative importance judgments in the study to be presented below are made in a prediction context, in which the commander (person) wants a mission to succeed (goal), and the subject considers causal elements of the situation (partition of the world) that are pertinent to this goal in light of the commander's plan of attack (behavior). Generally, successful communication of relative importance requires that the context in which relative importance was judged be evident to the recipient of the communication.

Definition of factors. Producing numerical measures of relative importance often involves considering tradeoffs: judging how much of an increase in one factor it would take to compensate for a decrease in another factor. These tradeoffs may be further related to the likelihood that these changes will happen, or to the amount of effort it would take to make them happen. Interpretation of the tradeoffs depends on unambiguous definitions of the factors. Researchers have observed a problem with people's consideration of such tradeoffs when judging relative importance: people are often insensitive to changes in the scale of measure- ment used in specifying the factors (Anderson and Zalinsky, 1988; Stewart and Ely, 1984). Successful communication of relative importance requires that the meaning of the factors be understood by sender and receiver, either because it is stated specifically or because it is already commonly understood.

Interdependencies among factors. Orderly consideration of tradeoffs is sometimes made difficult by interdependencies or intersubstitutabilities among the factors. For example, one factor may not be able to play any role unless another factor is at a particular level (e.g., it is not important to bring an umbrella on a trip unless it rains). If one has two umbrellas when it rains, neither one is very important unless the other is lost. A good measure of relative importance should allow people to express this sort of interdependency among the factors without confusion.

Measures of relative importance may be used to guide action when the following three conditions are met: the type of relative importance is clearly defined, the tradeoffs between factors are accessible, and the way to assign importance to interdependent factors is understood. Thus, in a resource allocation context the measure could be used to determine how much to allocate to each activity. In a scheduling context, the measure could tell in what order tasks should be done. In a decision situation, the measure could contribute to analysis of which course of action should be chosen.

Evaluating subjective measurement of relative importance. People's relative importance judgments are based on their intuitive conceptions of importance; hence they may not fit well with a given method. Therefore, we need to know what methods are best at eliciting and communicating people's judgments pertinent to each of the distinct formal models of relative importance. We also need to know how accurate people can be when using these methods to make these judgments. This knowledge would be important for redesigning the operations of a command center, for preparing curricula for the staff schools, and for designing decision aids that either receive or produce indications of relative importance.

What would it take for a method for eliciting subjective judgments of relative importance to be satisfactory? Several criteria may be identified:

1. **Reliability.** A person would judge that factors in the same situation have the same relative importances on different occasions.

2. **Intermeasure agreement.** The same person, applying two valid measures in the same situation, would produce similar profiles of relative importance. [However, if two measures have different definitions, as by referring to different goals, this criterion need not apply.]

3. **Interpersonal agreement.** Where the people share perspective, goals, and knowledge, their profiles using the same method in the same situation should be similar.

4. **Interpersonal agreement about interpretation of communications.** When people have knowledge of different parts of a situation, their interpretations of relative importance judgments should be the same. The recipient of a relative importance message should know what the sender means.

5. **Agreement with ground truth.** When analysis of the situation yields a true or justified measure of the factors' importance, the method should allow the knowledgeable person to produce relative importance judgments that are close to this measure.

6. **Internal consistency.** Relative importance judgments should conform to coherence constraints particular to the method. For example, probability is defined as a number between 0 and 1. The use of a probability change measure of relative importance (see below) should not produce a number outside this range.

7. **Consistency with factor definition.** The method should facilitate consideration of tradeoffs between the factors by being sensitive to changes in factor definition, yet it should not be oversensitive to such changes.

8. **Consistent treatment of factor interdependence.** The method should avoid confusion concerning the expression of knowledge about interdependence among the factors.

9. **Reliability in producing instruments for eliciting relative importance judgments in a particular situation.** It should be obvious what factors should be included, easy to avoid ambiguous factors and combinations of factors whose importance depends on each other, and obvious how to state each factor. The users' relative importance judgments should not be highly dependent on minor variations in how the factor is stated.

10. **Ease of use.** Users should find it easy to understand how to use the instrument. "Senders" should be motivated to use the instrument when needed, and "recipients" should find the relative importance judgments easy to read and interpret.

11. **Improvement of command and control.** If the relative importance measure is incorporated into a command and control system, performance should be improved.

The present study was designed to evaluate people's relative importance judgments on three of these criteria: intermeasure agreement, interpersonal agreement, and ease of use. It compares different methods for measuring relative importance in the context of predicting the success of a mission. During planning, commander and staff continually predict the success of the plans under consideration and use these judgments as the basis for changes. Therefore, judgments of the relative importance of different factors in determining a plan's success are central to the decision making function. The specific context for the relative importance judgments is: Given that there is little time to improve the plan, what changes in the plan are key?

Procedure

A 40-minute questionnaire was mailed to 520 army officers enrolled in the School of Corresponding Studies, the non-resident program at the Command and General Staff College, Fort Leavenworth, Kansas.

The questionnaire

The questionnaire presented two problems, each of which described a situation facing a commander, and presented the commander's plan. One dealt with an attack across a river by a

battalion led by LTC Rubel. The other was concerned with company commander CPT Smith's attempt to find a helicopter that has been heard to crash in a foggy valley. The problems, consisting of 2 pages of text plus a map of the situation, are in the appendix of Hamm (1991).

The respondent was asked to assess the probability of success of the commander's plan, to improve the plan, and to judge the probability of success of the improved plan. Then the respondent assessed the relative importance of 8 factors for the success of the commander's plan. The factors in the river problem were assumptions that were made in the problem: (a) the low strength of the enemy, (b) the availability of friendly artillery, (c) the fordability of the river, (d) that vehicles can pass over the ground on the far side of the river, (e) that the enemy does not occupy that ground, (f) that the enemy has not completed preparation of defensive positions, (g) that there will be no friendly air support, and (h) that there will be no enemy air support. The factors in the helicopter problem were either unknown aspects of the situation or aspects of the commander's hurried plan: (a) the fog in the valley where the crash occurred, (b) the actual location of the crash, (c) what the enemy is doing, (d) the decision to send one patrol rather than two, (e) the decision to send a small patrol, (f) the choice of the most experienced man to lead the patrol, (g) the choice of the most direct route to the estimated crash site, and (h) the decision not to request aerial reconnaissance before sending out the patrol.

Each respondent used two methods to rate the factors in the first problem (drawn from a total of five methods; see below). Then the second problem was read, its probability assessed, and the relative importance judged using the same two methods (in the same order).

After the second problem, respondents evaluated the two relative importance judgment methods they had used, as applied to each problem, with respect to seven issues: (a) ease of use, (b) confidence in accuracy of the method, (c) helpfulness for commander communicating to staff, (d) helpfulness for guiding the scheduling of planning tasks, (e) helpfulness for guiding the allocation of resources, (f) usefulness for explaining mission success, and (g) usefulness for explaining mission failure. Then they indicated which method they preferred and provided information about their background and experience.

Methods for judging relative importance

Five methods were used. The Yes or No, Ranking, Rating, and Probability Change methods each filled one page of the questionnaire. The Necessity or Sufficiency method filled two pages. Instructions for each method started with a general definition of importance, in the context of the particular problem. For the river problem, "The success of LTC Rubel's

attack plan depends on several important factors." The instructions continued as shown in the following paragraphs. The method for producing numerical scales for the responses, to facilitate comparison, is also described.

Yes or No. The simple, categorical response of this method imposes minimal demands for interfactor tradeoffs or recognition of interdependencies.

"Please consider the following list, and judge whether each factor is important in determining the success of the attack." [Each of the 8 factors given above was listed, with "Yes or No" placed to its left.] Responses were rescaled as follows: Yes = 1, No = 0.

Ranking. Ranking requires the user to attend to tradeoffs and adopt an approach for dealing with interdependencies.

"Please consider the following list, and rank the factors in order of their importance. That is, put a '1' by the <u>most important</u> factor, a '2' by the next most important, etc. Please do not use "ties": give a different rank number to each factor." [Each factor was listed, with a blank to its left.] Rescaling: ranks were reversed so that 8 = most important and 1 = least important.

Rating. This rating technique uses magnitude estimation instructions (Lodge, 1981; Stevens, 1975), which aim to give subjects' responses the quality of a ratio scale: if one factor is assigned a number twice as large as another, it means the factor is truly twice as important as the other. Rating forces the user to consider tradeoffs and interdependencies.

"Please consider the following list, and rate the factors according to their importance. Pick a fairly <u>unimportant</u> factor, and give it a score of '10'. Then give the other factors numbers that reflect how important they are, in relation to that first factor. Thus, you might give a very important factor a '200' because it is 20 times as important as the first, or you might give an unimportant factor a '5' because it is only half as important as the first factor. You can use fractions and can go as close to 0 as you want, or as high as you want." [Each factor was listed, with a blank to its left.]

Probability Change. Rather than asking judgments of "relative importance", this method focusses the user's attention on how the probability of success depends on the factors. By specifying that each factor changes while all others remain the same, the Probability Change method helps the user make tradeoffs and does not require expression of factor interdependencies.

"Please consider the following list, and estimate what the probability of attack success would be if the factor were to

change as described. Please give specific probabilities. Each of the changes should be considered independently, leaving all other features as described originally. Refer to the 'Plan Evaluation' question to review your estimate of the probability that the attack will succeed in taking Hill 434 by 1200 given current conditions and LTC Rubel's original plan."

To present each factor a specific alternative value, different from that given in the original problem presentation, was named. For example, the enemy strength was originally given as 50%, and the judgment for that factor was elicited using "What would be the probability of attack success if the enemy in fact were at 75% strength?". Rescaling: the absolute value of the change, between the original estimate of the plan's probability of success and the revised estimate given the changed factor, was taken. This was done because the effects of the changes in the factors varied: some changes made success more likely, some made it less likely.

Necessity or Sufficiency. This method addresses relative importance through the user's conception of the causal connections in a situation. Mackie (1974) proposed a formal vocabulary for this, in which factors could be characterized as "insufficient but necessary" elements of a higher level factor, which in turn is "unnecessary but sufficient" to bring about an effect. Because Mackie's concept is too complicated to explain in a questionnaire, judgments of simple necessity and sufficiency were used. While this method does not require the user to consider tradeoffs between factors, it explicitly addresses interdependencies.

"Please consider the following list, and indicate whether each is <u>necessary</u> for the success of the attack. In other words, could the attack succeed if the factor were not present? Also indicate whether each factor is <u>sufficient</u> for the success of the attack. In other words, could the attack possibly fail if the factor is present? Respond by circling Yes or No for each question. Note that some of the questions have been stated twice, using different words, to clarify their meanings."

The factors here were presented in a unique manner. For example, the enemy strength factor was presented as follows:

> "The enemy's low strength.
>
> Is the enemy's low strength (50%) necessary for the success of the attack? Would the attack probably fail if the enemy were stronger, say at 75% strength? Yes/No.
>
> Is the enemy's weakness sufficient for the success of the attack? Is the attack likely to succeed if the enemy is at only 50% strength? Yes/No."

Rescaling: Necessary = 1, Not Necessary = 0; Sufficient = 1, and Not Sufficient = 0. The scores for these two scales were summed.

Design

To assure that the five relative importance judgment techniques were used equally often, and to control for the order in which the two problems and the five methods were presented, a counterbalanced design was used. The factors were: 2 problem orders (helicopter problem then river problem, or the reverse) by 10 pairs of methods (all possible pairs of the five methods) by 2 method orders (within each pair). Half the cells in this design were dropped systematically, leaving 20 combinations.[1] A different questionnaire was constructed for each of these combinations.

Subjects were drawn from the active army, the reserves, and the national guard. Within these divisions, questionnaires (in blocks of 20) were sent to officers in each of six branches: armor, aviation, engineers, artillery, infantry, and signal corps.

Results

One hundred fifty three officers returned the questionnaire, for a return rate of 29.4%. No important differences were observed between active army, reserve, and national guard, nor between the branches, in return rate or other responses, and so these variables will not be further mentioned.

To produce comparable relative importance profiles across the 8 factors, for the five relative importance judgment methods, the means of the numerical measures (described above), across all the subjects who used each method (Ns are in Table 1, below), were normalized so that each factor's relative importance was on a scale from 0 to 1, and the scores for the 8

---

1. One of the 40 cells was "HktRkt", indicating that the helicopter problem (H) was first, judged with the Ranking (k) and then the Rating (t) methods, followed by the river problem (R) judged by the same methods in the same order. Without loss of information, we can call this "Hkt" ("Rkt" would indicate the river problem was presented first). Using this abbreviation scheme, with "y" = Yes or No method, "p" = Probability Change method, and "n" = Necessity or Sufficiency method, the 20 conditions in the study were: Hkt, Rtk, Hyk, Rky, Hkp, Rpk, Hnk, Rkn, Hty, Ryt, Hpt, Rtp, Htn, Rnt, Hpy, Ryp, Hyn, Rny, Hnp and Rpn.

factors summed to 1.[2] Results for the river problem are in Figure 1, below.

Internal coherence of responses

Respondents did not use the response scales strictly correctly. For example, logically speaking at most one factor can be both necessary and sufficient for the success of the mission, but the average respondent said that 3.4 of the 8 factors on the river problem were both necessary and sufficient. This may be due to the way the factors were described (see instructions above): in the attempt to clarify the meanings of "necessary" and "sufficient", expressions of degree of likelihood of success were used. These expressions invite respondents to pick a threshold other than 100% certainty when calling a factor "necessary" or "sufficient". As a result, it may be difficult to use the method to discover users' knowledge of interdependencies among the factors.

On every dimension there were a few subjects whose responses on the Probability Change dimension were in the opposite direction from the other respondents' responses. Most likely this was due to misinterpretation of the instructions (e.g., failure to refer to one's earlier estimate of the probability of mission success when judging the probability of success under changed conditions), although in some cases (see below) there was ambiguity about how the factor would affect the outcome of the mission.

Interpersonal variation in assignment of relative importance

There was individual variation in the importance judgments. A measure of this for a given problem is provided by the intercorrelations between different subjects' ratings of the 8 factors. Rather than computing all possible intercorrelations among more than 40 subjects, a sample of the intercorrelations is taken. All subjects who used a given method were placed in a ring and the correlation of each subject with each neighbor was calculated.[3] Subjects who gave equal responses to all 8 factors (e.g., some subjects using the Yes or No method said all factors were important) were dropped from this analysis.

The means of the samples of intersubject correlations for each method of judging relative importance are given in Columns 1 (Helicopter problem) and 3 (River problem) of Table 1. The number of subjects in the samples are in Columns 2 and 4,

---

2. Normalizing individually and then averaging produced nearly the identical results.

3. This was accomplished using the "lag" function of SPSS-PC, with a second pass to determine the correlation between the first and last subjects.

respectively. For the Ranking and Rating methods there was more intersubject agreement on the river problem, while for the Sufficiency judgments and hence for the Necessity or Sufficiency method there was more agreement on the helicopter problem. As a general estimate of the interpersonal agreement about relative importance using these methods, the unweighted mean of these mean correlations for the two problems was taken, producing the number in Column 5 of Table 1. (Row order in Table 1 is based on this measure.) Generally, the subjects agreed most on the Probability Change method, and least on the Sufficiency judgments. While there was relatively high intersubject agreement about the Rating and Ranking methods on the river problem (where the factors were assumptions upon which the plan was based), these methods had very low interpersonal agreement on the helicopter problem (where the factors were sources of uncertainty or aspects of the commander's plan). This shows that the interpersonal agreement of the relative importance measurement methods may be something that depends on the features of the particular situation being judged.

Table 1.

Intersubject correlations of relative importance judgments for eight situational factors

	Helicopter		River		Unweighted Mean of
Method	Mean r	n	Mean r	n	Mean r's
prob change	.21	52	.32	58	.265
necessity	.20	56	.25	59	.225
rating	.06	56	.33	57	.195
necess/suff	.23	57	.12	59	.175
ranking	.07	65	.26	65	.165
yes/no	.13	51	.17	43	.150
sufficiency	.18	56	.02	52	.100

Note: Mean r: the mean correlation between sampled pairs of subjects; n: the number of subjects in the sample, each subject being involved in two intersubject correlations; Unweighted mean of the mean r's: the mean of Columns 1 and 3, providing a more general estimate of intersubject agreement on each method.

When subjects' ratings of relative importance do not agree, it may be because they understand the factors differently, or because they disagree about how important the factors are. Some of the lack of correlation is due to different interpretations of the role of the factors in the situation. For example, with the

Probability Change method on the helicopter problem, 16 respondents thought that it would be an advantage if the fog lifted, changing the probability as much as +.65; 9 thought the fog made no difference; and 33 thought the plan would be less likely to succeed if the fog lifted, causing a change as large as -.60 in the probability of success. The Fog factor was exceptionally ambiguous, however, for there was greater agreement for the other factors in the helicopter problem and for each of the factors in the river problem. Because each subject read the problem, estimated the probability of success, and revised the commander's plan, it may be assumed that his or her interpretation of the factors was stable by the time he or she used the two methods to judge relative importance. Therefore the amount of disagreement in interpretation of the factors is presumably constant over all the methods, and so the between-method differences in interpersonal agreement in Table 1 may be attributed to differences in subjects' use of the methods to make relative importance judgments about the factors.

Similarity of relative importance ratings between methods

Figure 1 shows the relative importance profiles for the river problem, averaged across subjects for each method. Whether the attackers can ford the river, whether there will be friendly artillery support, and whether the enemy will have air support are deemed the most important factors. Friendly air support and whether the enemy has occupied positions before the target hill are considered least important. These relative importance profiles are fairly consistent across methods, with the exception of two factors: whether the river would be fordable (its normalized importance scores range from .11 in the Sufficiency measure to .26 in the Rating measure) and whether there would be friendly air support for the attack (scores range from .03 in the Necessity measure to .14 in the Sufficiency measure).

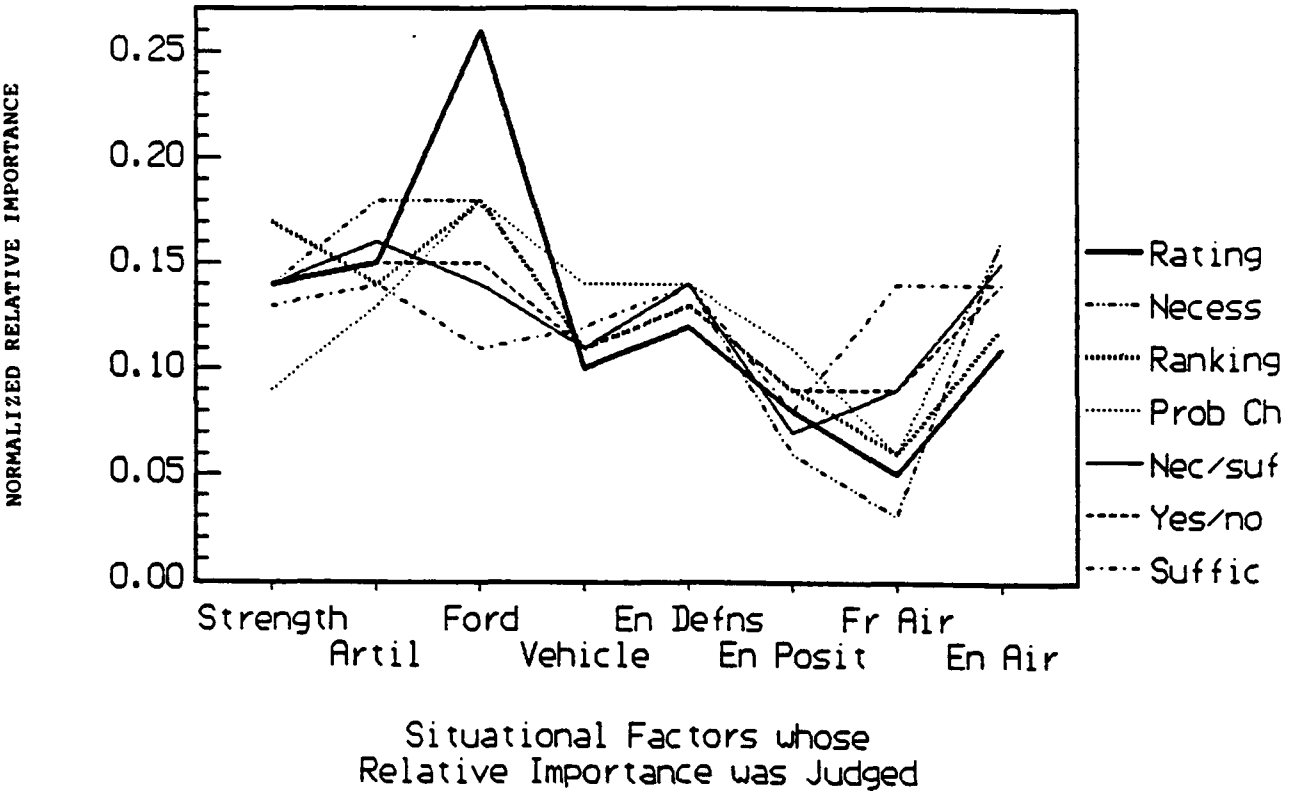


Figure 1. Profiles of relative importance produced by each method, and their mean.

The range of relative importance values varied between methods. The Yes or No method had the profiles with the narrowest range: all factors had mean normalized importance measures between .09 and .15. At the other extreme, the range of normalized importance scores for the Rating method was from .05 to .26. Therefore if it is desirable to highlight differences between factors, Ratings rather than Yes or No judgments should be used.

Intercorrelation among the profiles. Each subject produced a relative importance profile for the 8 factors of a problem, using each of two methods. The agreement between these profiles can be measured using correlation. Necessity and Sufficiency judgments are treated separately, as well as in combination (summed), in Table 2, which shows the mean intercorrelations, averaged over all subjects who did each pair of methods. The patterns for the two problems are similar: the Rating, Ranking,

and Yes or No methods were most strongly intercorrelated, Probability Change next, and the components of the Necessity or Sufficiency method the least.

Table 2.

Intermethod factor correlations

River Problem.

	Yes/No	Ranking	Rating	Prob Ch	Necess
Ranking	.66				
Rating	.78	.87			
Prob Chan	.58	.46	.50		
Necessary	.58	.58	.52	.52	
Sufficient	.36	.13	.03	.01	.30
Neces/Suf	.64	.45	.37	.34	

Helicopter Problem.

	Yes/No	Ranking	Rating	Prob Ch	Necess
Ranking	.69				
Rating	.70	.94			
Prob Chan	.62	.54	.52		
Necessary	.35	.27	.38	.12	
Sufficient	.21	.26	.07	.27	.33
Neces/Suf	.35	.31	.23	.23	

Note: Averages were taken with Fisher Z transformed correlations, then reverse transformed for display.

The Probability Change method differs from the others in that a specific alternative value is named for every factor. Thus, the respondent compares the importance of these specified changes (from the given value on the factor to the alternative value) to each other, while in each of the other methods the comparison is between the given values (perhaps interpreted with reference to each individual's default value). Because of this difference, there is no a priori reason that there should be any similarity between the Probability Change method's relative importance profiles and the profiles of the other methods. Rather, similarity depends on whether the person who specifies the changes in the factors, for the Probability Change method, names alternatives similar to those that the respondents spontaneously think of when using the other methods.

Respondents' preferences for relative importance assessment methods

Each subject used two methods for expressing relative importance, and said which one was better. Table 3 shows how many subjects preferred each method, in each pair comparison. Ranking was the most chosen, then Probability Change and Yes or No, then Necessity or Sufficiency and Rating. The preferences were far from unanimous: 42 respondents preferred the Ranking method while 19 preferred the method it was paired with; at the other extreme, 17 preferred the Rating method while 35 preferred its competitor. The relative strengths of preference are not fully transitive. For example, the Yes or No method was preferred over the Necessity or Sufficiency method, 13 to 2; Necessity or Sufficiency beats Rating 7 to 4; and yet Yes or No only beat Rating by 8 to 7.

Table 3.

Pairwise Preferences for Relative Importance Methods

Preferred Method	Non-Preferred Method: Ranking	Prob Change	Yes/No	Rating	Nec/Suff	Total Preferred	Proportion
Ranking	~	5	10	11	16	42	.69
Prob Change	6	~	10	9	8	33	.65
Yes/No	5	5	~	8	13	31	.52
Rating	5	1	7	~	4	17	.33
Nec/Suff	3	7	2	7	~	19	.32
Total Non-preferred	19	18	29	35	41	142	

Respondents' evaluations of methods

Respondents rated each method, as applied to each problem, on seven attributes (listed in Procedure section, above), using scales whose values range from 1 to 7. The mean ratings (collapsed across problem) are shown in Figure 2. The Ranking and Probability Change methods received higher average ratings than all others, except that the Yes or No method was judged easiest.

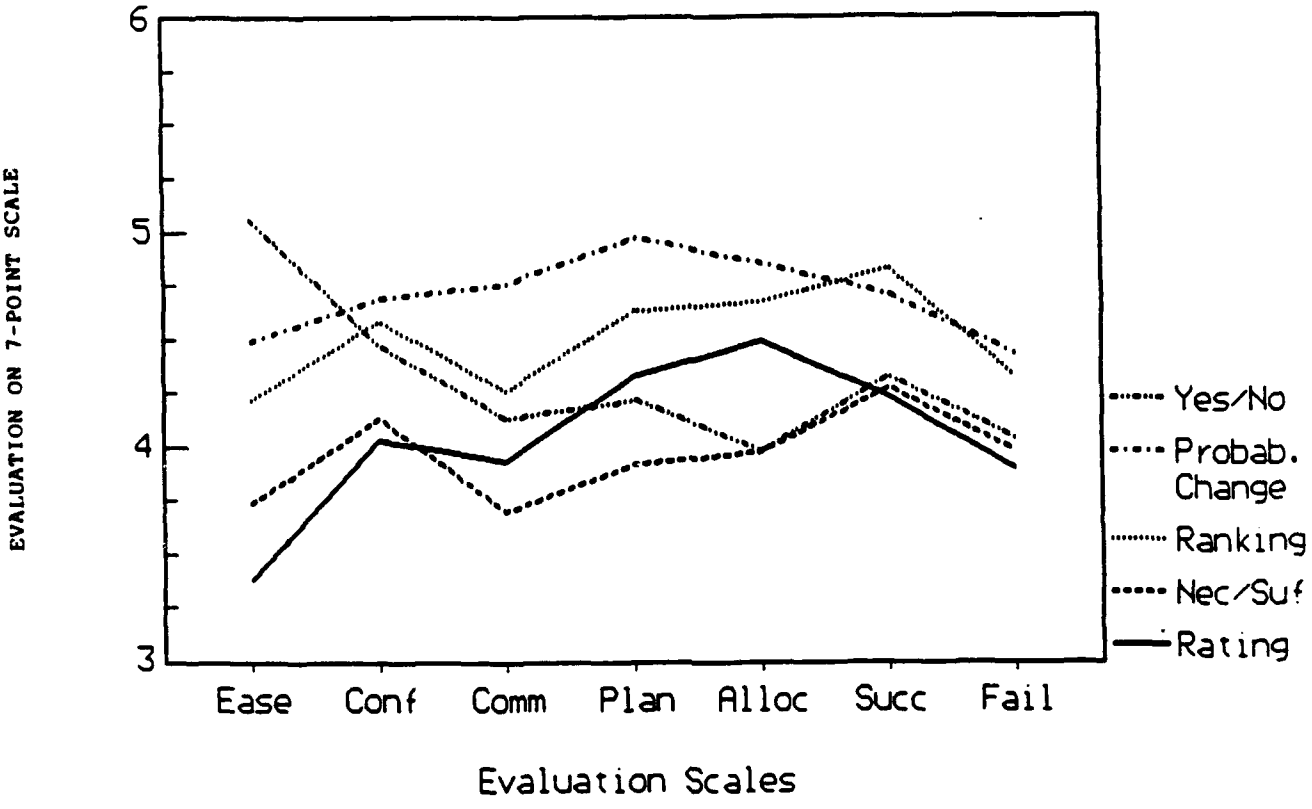


Figure 2. Mean evaluative ratings for each relative importance assessment method.

Discussion

The study investigated five alternative methods for measuring the relative importance of situational factors for predicting mission success in a planning context. It was designed to evaluate the methods using three of the eleven possible criteria presented above.

Interpersonal agreement

The respondents had very different opinions about the relative importance of the factors in the situations, and about the methods they used to express their relative importance judgments. Intercorrelations among subjects using the same methods were relatively low, and the judgments of relative importance of the factors had large standard deviations. In

paired comparisons, at least 32% of the respondents favored each of the relative importance methods. Finally, the respondents gave each method nearly the whole range, from 1 to 7, on each evaluation question.

Perhaps this variability is due to particular characteristics of the study. Questionnaires completed in private do not inspire the most careful attention. The problems may not be realistic because they offer a brief description of a situation about which a real staff officer would know and care much more. A second explanation is that people are unfamiliar with the techniques for judging relative importance, and therefore use them unreliably. Presumably familiarity with the methods would make their judgments agree more. Even though each method might produce different profiles, the subjects would use the same methods in similar ways. Third, the instructions may be inadequate because they do not specify how the subjects should express their knowledge of interdependencies. Fourth, the presentation of the methods may be inadequate because the factors are not defined specifically enough to let the subjects make unambiguous tradeoffs among the factors. Fifth, the variation may reflect legitimate differences of opinion. A sixth explanation, however, is that the source of the respondents' disagreements is their fundamental confusion about the meaning of relative importance (Goldstein and Beattie, in press; Hamm, 1990b). Familiarity with the methods, alone, would not clear this up; it would require that clarifying distinctions be made and promulgated throughout the organization.

Intermethod agreement

Comparison of the relative importance profiles that different methods produce gives information on three issues. First, if the different methods produce similar profiles it validates the general concept of relative importance. Second, when methods have systematic disagreements it suggests that there is an important distinction between them, and that analysis is needed to give guidance on which method to use in which type of context. Third, methods that produce similar profiles may probably be substituted for each other in response to user preference.

The degree of intermethod agreement was not extremely high. The Rating, Ranking, and Yes or No methods were most strongly intercorrelated, suggesting safest intersubstitution. Among these, the Ranking method was best liked although the Yes or No method was easiest.

Probability Change and Necessity or Sufficiency showed lower agreement with the other methods. This suggests that measuring relative importance using the vocabulary of probability or of necessity or sufficiency induces people to think about the problem differently. There may be special advantages to using

these methods, but the resulting profiles should not simply be labeled "relative importance" or confusion may result.

Ease of use

Subjects evaluated the relative importance judgment methods using seven subjective rating scales. The Probability Change and Ranking methods were generally the most positively rated, although the Yes or No method was called the easiest to use. These three methods were also most often preferred.

These subjective rating scales have two problems. First, the ratings are highly intercorrelated, so that it is not possible to draw firm conclusions about which methods are better for the various purposes (e.g., for a commander giving the staff guidance in deciding in what order to undertake various planning tasks, as contrasted with explaining why a mission failed). Second, there is the possibility that the high intercorrelations are due to a halo effect in which the ratings are based on the subject's general evaluation of the method rather than on the specific evaluation criteria. This problem is inherent in all uses of subjective rating scales to elicit multi-dimensional evaluations of equipment, decision aids, etc. Research measuring specific objective components of "ease", such as speed of relative importance judgments or training time to an accuracy criterion, would be needed to address more adequately the question of which methods are easy to use.

Conclusion

The present study offers a preliminary exploration of the use of relative importance assessment in military command and control. It produced information concerning three of eleven criteria for evaluating methods for eliciting relative importance judgments: intermeasure agreement, interpersonal agreement, and ease of use (with some observations concerning internal consistency for the Necessity or Sufficiency method). It covered just a sample of the possible methods for measuring relative importance, and applied them only in the causal prediction context (see Hamm, 1990b, for four other contexts).

Some of the methods give the subject more guidance in thinking about the factors than others. Other methods are perceived as requiring more work (see Figure 2). Table 4 shows the application of these two distinctions to the five methods. These distinctions could affect: (a) how difficult it is to produce elicitation forms and to use the methods, (b) people's preferences for the methods, and (c) the accuracy and interpersonal agreement of the methods. This information can partially guide the developer of a decision aid wishing to include subjective assessments of relative importance weights.

Table 4.

Differences between methods in guidance given and work required

Amount of work required of the respondent	Amount of guidance given to the respondent. Low	High
Low	Yes/No, Ranking	Probability Change
High	Rating	Necessity/ Sufficiency

The specific findings of the study can also help those who wish to make use of relative importance judgments. The Probability Change method had relatively high intersubject agreement, on both problems, and hence is a candidate for a general method for expressing relative importance when interpersonal agreement is strongly needed. The correlations among the profiles indicate that the Rating, Ranking, and Yes or No methods can most safely be substituted for each other. According to the subjective evaluations, the Ranking, Probability Change, and Yes or No methods would be most acceptable to users. Note that the acceptability and intersubstitutability of the Ranking method will depend on the number of factors being ranked (n): as n increases, the difficulty of Ranking increases rapidly (the number of comparison judgments goes up proportional to n^2), while the difficulty of the other methods increases more slowly (the number of judgments goes up proportional to n).

The study revealed some problems with relative importance judgments. When used by isolated subjects in a briefly described context, agreement between subjects and between methods is not very high and subjects do not always use terms in accordance with their definitions. At this stage, we do not know the theoretical limits of people's ability to make relative importance judgments. Research is required to answer the following types of question.

If people who have extensive experience in the situation were to use the methods to communicate about the relative importance of meaningful factors, would the judgments have better qualities in terms of the criteria measured here (interpersonal and intermethod agreement), as well as the others

listed above? Is it possible to train people to make better relative importance judgments? Training might focus on: recognizing which type of relative importance context is appropriate and conveying this to others; selecting the factors to judge, including minimizing interdependencies among factors without making the factor definitions too abstract; correctly considering tradeoffs among dimensions; conveying to others the factor definitions on which one bases one's relative importance judgments (in sufficient detail to support their interpretation of one's relative importance measures in terms of specific tradeoffs between the factors); recognizing interdependencies among factors, using one's chosen relative importance measurement method to deal with such interdependencies, and assuring that others understand that one is doing so.

If the methods used in this study are employed correctly by trained users in meaningful environments, how much work will it require and how accurate can the judgments be? Will judgment of relative importance using the best methods prove useful or should we rely instead on more specific concepts that serve the same function (e.g., judgments in terms of formal models of resource allocation, judgments of necessity or sufficiency or of the more complicated concepts in Mackie's (1974) framework, or judgments of the impacts of specified situational changes on probabilities of specified outcomes)?

The answers to these questions could be useful in assuring that communication concerning the importance of various factors in the situation will be accomplished with minimal confusion.

References

Andriole, S.J. (1987). TACPLAN -- An intelligent aid for army tactical planning. In S.J. Andriole (Ed.), Artificial Intelligence and National Defense: Applications to C^3I and Beyond (pp 143-153). Washington, D.C.: AFCEA International Press.

CGSC (1988). A guide to the estimate of the situation (Student Text 100-9). Fort Leavenworth, KS: U.S. Army Command and General Staff College.

Coates, E.N., and McCourt, A.W. (1976). A Questionnaire-based Analysis of Order-of-battle Elements (ARI Technical Paper 271). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A021 956)

Cook, R.L., and Stewart, T.R. (1975). A comparison of seven methods for obtaining subjective descriptions of judgmental policy. Organizational Behavior and Human Performance, 13, 31-45.

Goldstein, W.M., and Beattie, J. (in preparation). Judgments of relative importance in decision making: The importance of interpretation and the interpretation of importance. In D.R. Brown and J.E.K. Smith (Eds.), Frontiers in Mathematical Psychology. New York: Springer-Verlag.

Hamm, R.M. (1990a). Comparison of five methods for expressing the relative importance of causal factors in predicting mission success. In Proceedings, Seventh Annual Workshop on Command and Control Decision Aiding. Dayton, Ohio: Air Force Institute of Technology, Wright-Patterson Air Force Base.

Hamm, R.M. (1990b). Expressions of relative importance. In I.B. Turksen (Ed.), Proceedings of NAFIPS'90 Conference, Volume II (pp. 382-385). Toronto: North American Fuzzy Information Processing Society/University of Toronto.

Hamm, R.M. (1991). Judgments of probability and relative importance in a military decision scenario: The influence of variations in relevant and irrelevant causal factors. (ARI Technical Report 931). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A239 935)

Jaccard, J., and Sheng, D. (1984). A comparison of six methods for assessing the importance of perceived consequences in behavioral decisions: Applications from attitude research. Journal of Experimental Social Psychology, 20, 1-28.

Lodge, M. (1981). Magnitude Scaling: Quantitative Measurement of Opinions. Beverly Hills, CA: Sage Publications.

Mackie, J.L. (1974). The Cement of the Universe: A Study of Causation. Oxford: Clarendon Press.

McNemar, Q. (1962). Psychological Statistics. New York: Wiley.

Reilly, B.A., and Doherty, M.E. (1989). A note on the assessment of self-insight in judgment research. Organizational Behavior and Human Decision Processes, 44, 123-131.

Rosenthal, R., and Rosnow, R.L. (1984). Essentials of Behavioral Research: Methods and Data Analysis. New York: McGraw-Hill Book Company.

Shanteau, J. (1980). The Concept of Weight in Judgment and Decision Making: A Review and Some Unifying Proposals (Report No. 228). Boulder, CO: Center for Research on Judgment and Policy, University of Colorado.

Stevens, S.S. (1975). Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. NY: John Wiley.

Stewart, T.R., and Ely, D.W. (1984). Range sensitivity: A necessary condition and a test for the validity of weights (Report No. 270). Boulder, CO: Center for Research on Judgment and Policy, University of Colorado.

U.S. Army. (1954). Staff Officers' Field Manual: Staff Organization and Procedure (Field Manual 101-5). Washington, D.C.: Department of the Army.

U.S. Army. (1986). Operations (Field Manual 100-5). Washington, D.C.: Department of the Army.

U.S. Army. (1987). Leadership and Command at Senior Levels (Field Manual 22-103). Washington, D.C.: Department of the Army.